Komp'juternaja Lingvistika i Intellektual'nye Tehnologii 2015 vol.2 N14, pages 65-75

A supervised approach for SentiRuEval task on sentiment analysis of tweets about telecom and financial companies

Tutubalina E., Zagulova M., Ivanov V., Malykh V.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

This paper describes a supervised approach for solving a task on sentiment analysis of tweets about banks and telecom operators. The task was articulated as a separate track in the Sentiment Evaluation for Russian (SentiRuEval-2015) initiative. The approach we proposed and evaluated is based on a Support Vector Machine model that classifies sentiment polarities of tweets. The set of features includes term frequency features, twitter-specific features and lexicon-based features. Given a domain, two types of sentiment lexicons were generated for feature extraction: (i) manually created lexicons, constructed from Pros and Cons reviews; (ii) automatically generated lexicons, based on pointwise mutual information between unigrams in a training set. In the paper we provide results of our method and compare them to results of other teams participated in the track. We achieved 35.2% of macro-averaged F-measure for banks and 44.77% for tweets about telecom operators. The method described in the paper is ranked second and fourth among 7 and 9 teams, respectively. The best SVM setting after tuning parameters of the classifier and error analysis with common types of errors are also presented in this paper.

Keywords

Sentiment analysis, Sentirueval, Social media, Tweet sentiment classification, Twitter